

# **ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI TINGKAT KESEJAHTERAAN KELUARGA MENGGUNAKAN ALGORITMA C4.5**



**Disusun sebagai salah satu syarat menyelesaikan Jenjang Strata I  
Pada Jurusan Informatika Fakultas Komunikasi dan Informatika**

**Oleh:**

**KURNIA IKA PERMATASARI**

**L 200 120 131**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
UNIVERSITAS MUHAMMADIYAH SURAKARTA  
2016**

**HALAMAN PERSETUJUAN**

**ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI TINGKAT  
KESEJAHTERAAN KELUARGA MENGGUNAKAN ALGORITMA C4.5**

**PUBLIKASI ILMIAH**

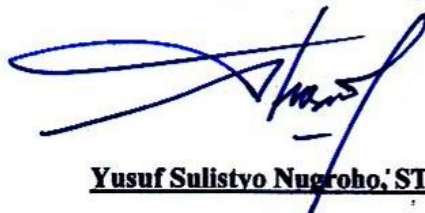
oleh:

**KURNIA IKA PERMATASARI**

**L 200 120 131**

Telah diperiksa dan disetujui untuk diuji oleh:

Dosen Pembimbing



**Yusuf Sulisty Nugroho, ST., M.Eng.**

**NIK.1197**

## HALAMAN PENGESAHAN

### ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI TINGKAT KESEJAHTERAAN KELUARGA MENGGUNAKAN ALGORITMA C4.5

OLEH

KURNIA IKA PERMATASARI

L 200 120 131

Telah dipertahankan di depan Dewan Penguji  
Fakultas Komunikasi dan Informatika  
Universitas Muhammadiyah Surakarta  
Pada hari Jum'at, 24 Juni 2016  
dan dinyatakan telah memenuhi syarat


Dewan Penguji:

1. Yusuf Sulisty N., S.T., M.Eng.  
(Ketua Dewan Penguji)
2. Dedi Gunawan, S.T., M.Sc.  
(Anggota I Dewan Penguji)
3. Fatah Yasin, S.T., M.T.  
(Anggota II Dewan Penguji)

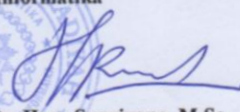
(.....)  
(.....)  
(.....)

Mengetahui,

Dekan  
Fakultas Komunikasi dan Informatika

  
Husni Thamrin, S.T., M.T., Ph.D.  
NIK : 706

Ketua Program Studi  
Informatika

  
Dr. Heru Suprivono, M.Sc.  
NIK : 970

## **PERNYATAAN**

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali secara tertulis diacu dalam naskah dan disebutkan dalam daftar pustaka.

Apabila kelak terbukti ada ketidakbenaran dalam pernyataan saya di atas, maka akan saya pertanggungjawabkan sepenuhnya.

**Surakarta, 24 Juni 2016**

Penulis



**KURNIA IKA PERMATASARI**

**L 200 120 131**



**UNIVERSITAS MUHAMMADIYAH SURAKARTA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
PROGRAM STUDI INFORMATIKA**

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448  
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: [informatika@ums.ac.id](mailto:informatika@ums.ac.id)

---

**SURAT KETERANGAN LULUS PLAGIASI**

**012/A.3-IL3/INF-FKI/VII/2016**

Assalamu'alaikum Wr. Wb

Biro Tugas Akhir Program Studi Informatika menerangkan bahwa :

Nama : KURNIA IKA PERMATASARI  
NIM : L200120131  
Judul : ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI TINGKAT  
KESEJAHTERAAN KELUARGA MENGGUNAKAN ALGORITMA  
C4.5  
Program Studi : Informatika  
Status : **Lulus**

Adalah benar-benar sudah lulus pengecekan plagiasi dari Naskah Publikasi Tugas Akhir,  
dengan menggunakan aplikasi Turnitin.

Demikian surat keterangan ini dibuat agar dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Wr. Wb

Surakarta, 25 Juli 2016

Biro Tugas Akhir Informatika

**Endang Wahyu Pamungkas, S.Kom., M.Kom.**

preferences

previous paper next paper



Originality Report

Processed on: 23-Jul-2016 08:43 WIB  
ID: 691184319  
Word Count: 2907  
Submitted: 1

# ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI TINGKAT KESEJAHTERAAN KELUARGA MENGGUNAKAN ALGORITMA C4.5 Abstrak

By Kurnia Ika Permatasari

Similarity Index	Similarity by Source
30%	Internet Sources: 19%
	Publications: 3%
	Student Papers: 22%

Document Viewer

exclude quoted exclude bibliography exclude small matches

mode: show highest matches together

ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI TINGKAT KESEJAHTERAAN KELUARGA MENGGUNAKAN ALGORITMA C4.5 Abstrak

**Kemiskinan merupakan fenomena sosial, tidak hanya di negara-negara berkembang, tetapi juga negara-negara maju.**

6

Berdasarkan data kemiskinan Badan Pusat Statistik 2015 pada periode 2009 - Maret 2015 kemiskinan penduduk Indonesia mencapai angka 28,59 juta jiwa. Hal ini membuktikan bahwa masih banyak faktor-faktor yang harus dibenahi dari setiap keluarga yang menyebabkan kemiskinan masih ada. Penelitian ini dilakukan untuk membantu keluarga dan pemerintah dalam memprediksi tingkat kesejahteraan keluarga dari segi ekonomi serta

**untuk mengetahui faktor-faktor yang paling berpengaruh terhadap**

17

tingkat kesejahteraannya melalui sebuah aplikasi. Algoritma C4.5 digunakan untuk mencari nilai gain ratio terbesar dari atribut-atribut yang digunakan dalam penelitian ini, seperti

**padi-padian, umbi-umbian, ikan/udang/ cumi/ kerang, daging, telur dan susu, sayur-sayuran, kacang-kacangan, buah-buahan, minyak dan kelapa, bahan minuman, bumbu-bumbuan, konsumsi lainnya, makanan dan minuman jadi, rokok, perumahan dan fasilitas rumah tangga,**

1

1 7% match (Internet from 08-Jun-2016)  
<http://repository.ipb.ac.id>

2 5% match (student papers from 21-Apr-2016)  
Class Publikasi Wisuda Juni  
Assignment Publikasi Wisuda Juni  
Paper ID: [662961257](#)

3 2% match (student papers from 04-Feb-2014)  
Class publikasi maret 2014  
Assignment publikasi\_maret\_2014  
Paper ID: [393332890](#)

4 2% match (Internet from 30-May-2016)  
<http://jurnal-tip.net>

5 2% match (student papers from 25-Jun-2015)  
Class publikasi  
Assignment publikasi september 2015  
Paper ID: [552617733](#)

6 1% match (Internet from 01-Apr-2011)



# ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI TINGKAT KESEJAHTERAAN KELUARGA MENGGUNAKAN ALGORITMA C4.5

## Abstrak

Kemiskinan merupakan fenomena sosial, tidak hanya di negara-negara berkembang, tetapi juga negara-negara maju. Berdasarkan data kemiskinan Badan Pusat Statistik 2015 pada periode 2009 - Maret 2015 kemiskinan penduduk Indonesia mencapai angka 28,59 juta jiwa. Hal ini membuktikan bahwa masih banyak faktor-faktor yang harus dibenahi dari setiap keluarga yang menyebabkan kemiskinan masih ada. Penelitian ini dilakukan untuk membantu keluarga dan pemerintah dalam memprediksi tingkat kesejahteraan keluarga dari segi ekonomi serta untuk mengetahui faktor-faktor yang paling berpengaruh terhadap tingkat kesejahteraannya melalui sebuah aplikasi. Algoritma C4.5 digunakan untuk mencari nilai *gain ratio* terbesar dari atribut-atribut yang digunakan dalam penelitian ini, seperti padi-padian, umbi-umbian, ikan/udang/cumi/ kerang, daging, telur dan susu, sayur-sayuran, kacang-kacangan, buah-buahan, minyak dan kelapa, bahan minuman, bumbu-bumbuan, konsumsi lainnya, makanan dan minuman jadi, rokok, perumahan dan fasilitas rumah tangga, aneka barang dan jasa, pakaian, alas kaki dan tutup kepala, barang tahan lama, pajak, pungutan dan asuransi, keperluan pesta/upacara/kenduri, dan pendapatan. Hasil penelitian yaitu berupa sebuah aplikasi yang digunakan untuk memprediksi tingkat kesejahteraan keluarga dan menyajikan informasi berupa faktor-faktor yang mempengaruhinya sesuai dengan atribut yang ada. Sehingga dapat menjadi pertimbangan bagi keluarga dan pemerintah di masa mendatang.

**Kata Kunci:** BPS (Badan Pusat Statistik), Data Mining, Kesejahteraan keluarga, Algoritma C4.5.

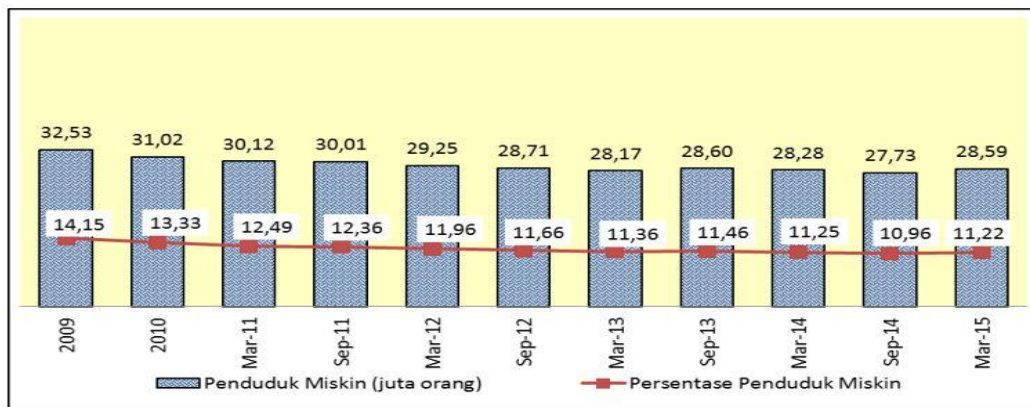
## Abstract

Poverty is a social phenomenon, not only in developing countries but also in developed countries. Based on data from the Bureau of Statistics Center 2015, in the period of 2009 until - March 2015 poverty among Indonesia's population reached 28.59 million. It proves that there are still many factors that must be addressed from every family that causes poverty. The objective of this study is to help families and governments in predicting welfare level in terms of economic as well as to determine the most influence factor their welfare levels using an application. C4.5 algorithm is used to find the highest gain ratio of the attributes used in this study, such as grains, tubers, fish / shrimp / squid / shellfish, meat, eggs and milk, vegetables, nuts, fruits, oils and coconut, beverage ingredients, spices, other consumables, food and beverages, cigarettes, housing and household facilities, miscellaneous goods and services, clothing, footwear and headgear, durable goods, tax levies and insurance, the purpose of the party / ceremony / festivity, and income. The result of this study is an application used to predict the welfare level of the family and provide the information of influencing factors in accordance with the attributes. So that may be a consideration for families and governments in their future economic planning.

**Keywords:** BPS (Bureau of Statistics Center), Data Mining, Welfare level, C4.5 algorithm.

## 1. PENDAHULUAN

Kemiskinan merupakan fenomena sosial, tidak hanya di negara-negara berkembang, tetapi juga di negara-negara maju. Gambar 1 menunjukkan data kemiskinan Badan Pusat Statistik 2015 dari periode 2009 - Maret 2015 yang menunjukkan bahwa tingkat kemiskinan di Indonesia mengalami penurunan kecuali pada September 2013 dan Maret 2015 (BPS,2015).



Sumber: Diolah dari data Survei Sosial Ekonomi Nasional (Susenas)

Gambar 1. Jumlah dan Persentase Penduduk miskin 2009-2015

Berdasarkan konsep kemiskinan BPS, kemiskinan dipandang sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran. Jadi keluarga yang memiliki rata-rata perkapita perbulan dibawah garis kemiskinan termasuk kelompok tidak sejahtera. Garis Kemiskinan (GK) merupakan penjumlahan dari Garis Kemiskinan Makanan (GKM) dan Garis Kemiskinan Non Makanan (GKNM) (BPS Kabupaten Ngawi, 2015). Berdasarkan data - data pada gambar 1 dapat ditunjukkan bahwa masih banyak keluarga yang memiliki pendapatan di bawah Garis Kemiskinan (GK). Maka perlu adanya faktor yang harus dibenahi dari setiap keluarga yang menyebabkan kemiskinan di Indonesia meningkat, karena keluarga merupakan unit terkecil dalam masyarakat yang memberikan pengaruh besar terhadap kinerja pembangunan dalam mendukung program-program pemerintah yang ada. Melalui keluarga yang sejahtera ini, dapat menjadikan kehidupan bangsa lebih tentram, aman, harmonis dan damai. Dengan demikian, kesejahteraan keluarga menjadi salah satu tolak ukur dalam pembangunan dengan program-program yang dibentuk oleh pemerintah. Pada tabel 1 menunjukkan indikator kemiskinan menurut Kabupaten/Kota 2014.

Tabel 1. Indikator Kemiskinan

KABUPATEN/KOTA		2014				
		GK	Jumlah (000)	%	P1	P2
3501	Pacitan	220.810	88.94	16.18	2.23	0.49
3502	Ponorogo	247.368	99.86	11.53	1.49	0.32
3503	Trenggalek	250.666	90.04	13.10	1.58	0.32
3504	Tulungagung	277.707	88.99	8.75	1.03	0.21
3505	Blitar	244.382	116.72	10.22	1.50	0.35
3506	Kediri	251.547	196.78	12.77	1.75	0.39
3507	Malang	254.380	280.31	11.07	1.51	0.39
3508	Lumajang	234.728	120.69	11.75	1.04	0.16
3509	Jember	267.962	270.40	11.28	1.11	0.20
3510	Banyuwangi	285.004	147.70	9.29	1.38	0.33



<b>3511</b>	Bondowoso	299.819	111.88	14.76	1.84	0.45
<b>3512</b>	Situbondo	246.483	87.67	13.15	1.87	0.46
<b>3513</b>	Probolinggo	340.539	231.92	20.44	2.99	0.73
<b>3514</b>	Pasuruan	283.327	170.74	10.86	1.59	0.41
<b>3515</b>	Sidoarjo	346.538	133.83	6.40	0.83	0.20
<b>3516</b>	Mojokerto	293.609	113.33	10.56	1.19	0.26
<b>3517</b>	Jombang	301.162	133.50	10.80	1.20	0.25
<b>3518</b>	Nganjuk	308.506	136.50	13.14	1.75	0.39
<b>3519</b>	Madiun	265.310	81.20	12.04	1.37	0.27
<b>3520</b>	Magetan	262.069	73.97	11.80	0.94	0.15
<b>3521</b>	<b>Ngawi</b>	<b>240.780</b>	<b>123.18</b>	<b>14.88</b>	<b>1.64</b>	<b>0.32</b>
<b>3522</b>	Bojonegoro	272.886	190.88	15.48	2.20	0.54
<b>3523</b>	Tuban	265.659	191.13	16.64	2.53	0.64
<b>3524</b>	Lamongan	289.403	186.12	15.68	1.78	0.36
<b>3525</b>	Gresik	348.888	166.95	13.41	2.20	0.65
<b>3526</b>	Bangkalan	305.174	212.16	22.38	3.47	0.88
<b>3527</b>	Sampang	272.900	239.60	25.80	3.88	0.87
<b>3528</b>	Pamekasan	266.953	148.76	17.74	2.51	0.63
<b>3529</b>	Sumenep	270.890	218.86	20.49	3.32	0.81
<b>3571</b>	Kota Kediri	366.788	22.13	7.95	1.43	0.43
<b>3572</b>	Kota Blitar	319.177	9.81	7.15	0.88	0.18
<b>3573</b>	Kota Malang	381.400	40.64	4.80	0.43	0.06
<b>3574</b>	Kota Probolinggo	574.105	19.03	8.37	2.98	0.75
<b>3575</b>	Kota Pasuruan	328.648	14.21	7.34	1.05	0.24
<b>3576</b>	Kota Mojokerto	328.250	8.02	6.42	0.73	0.13
<b>3577</b>	Kota Madiun	338.609	8.48	4.86	0.76	0.20
<b>3578</b>	Kota Surabaya	393.151	164.36	5.79	0.60	0.11
<b>3579</b>	Kota Batu	355.317	9.14	4.59	0.43	0.06
<b>35</b>	<b>JAWA TIMUR</b>	<b>289.945</b>	<b>4.748.42</b>	<b>12.28</b>	<b>1.86</b>	<b>0.45</b>

Sumber : Badan Pusat Statistik 2014

Dalam penelitian ini akan dilakukan proses *data mining* guna untuk mengetahui tingkat kesejahteraan keluarga dan faktor yang paling mempengaruhinya di kabupaten Ngawi. Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basisdata. Data dalam suatu *database* dapat diekstraksi dan dikenali polanya untuk menghasilkan informasi yang sangat berguna (Nugroho, 2014). Metode pohon keputusan menggunakan algoritma C4.5 digunakan dalam penelitian, metode ini merupakan sebuah metode untuk menghasilkan pohon keputusan yang dapat diinterpretasikan untuk melakukan suatu prediksi, memiliki tingkat akurasi yang dapat diterima, efisien dalam menangani atribut bertipe diskret dan numerik (Kamagi dan Hasun, 2014).

Nugroho dan Setyawan (2014) dalam penelitiannya menyatakan bahwa dalam dunia pendidikan data yang berlimpah dan berkelanjutan bisa dimanfaatkan untuk data mining dalam rangka pengelolaan yang lebih baik dan pelaksanaan pembelajaran yang lebih efektif. Metode *decision tree* dengan algoritma C4.5 digunakan dalam penelitian ini sebagai sumber informasi strategis bagi fakultas untuk mengklasifikasi masa studi mahasiswa dengan menggunakan teknik data mining. Atribut yang digunakan terdiri dari jurusan sekolah, jenis kelamin, asal sekolah, rerata jumlah SKS per semester, dan peran menjadi asisten. Hasil penelitian menunjukkan bahwa variabel yang paling tinggi pengaruhnya terhadap masa studi mahasiswa adalah rerata SKS per semester. Hasil penelitian mengindikasikan bahwa variabel yang perlu digunakan sebagai pertimbangan bagi fakultas untuk memperoleh tingkat masa studi yang efektif adalah rerata SKS yang diambil oleh mahasiswa.

Berdasarkan dari telaah penelitian sebelumnya tentang analisis faktor-faktor yang mempengaruhi tingkat kesejahteraan keluarga menggunakan algoritma C4.5, bahwa *gain ratio* digunakan untuk mencari nilai terbesar dari semua atribut yang ada. Data sampel yang digunakan yaitu pengeluaran ekonomi keluarga di Kabupaten Ngawi. Penelitian ini dilakukan untuk membantu keluarga dan pemerintah dalam memprediksi tingkat kesejahteraan keluarga dari segi ekonomi serta untuk mengetahui faktor-faktor yang paling berpengaruh terhadap tingkat kesejahteraannya melalui sebuah aplikasi berbasis web. Sehingga dapat membantu keluarga dan pemerintah di masa mendatang.

## 2. METODE

### 2.1 Analisis Data Mining

Pada penelitian ini dilakukan untuk mencari nilai *gain ratio* terbesar pada masing – masing atribut dari data pengeluaran ekonomi keluarga selama seminggu, sebulan dan setahun untuk memprediksi tingkat kesejahteraan keluarga berdasarkan atribut-atribut yang telah ditentukan. Analisis data mining memiliki tahapan-tahapan yang harus dilakukan untuk mencapai hasil yang sesuai dengan tujuan.

#### 2.1.1 Pengumpulan Data

Data-data yang digunakan dalam penelitian ini yaitu data pengeluaran dan pendapatan ekonomi keluarga di Kabupaten Ngawi sebagai data *training*. Dengan menggunakan metode Slovin (Nugroho dan Haryati, 2015), data *training* diambil sebanyak 399 data sampel keluarga dari total 251.790 keluarga yang berada di Kabupaten Ngawi dari survei secara random, yang dapat dilihat pada persamaan 1. Disamping itu juga menggunakan data *testing* yang diperoleh dari sampel *user*.

$$n = \frac{N}{1 + Ne^2} \dots\dots\dots (1)$$

Keterangan :

n = jumlah sampel

$N$  = jumlah keseluruhan data / populasi

$e$  = galat kesalahan (ditentukan sebesar 5%)

### 2.1.1.1 Data Training

Data *training* merupakan data yang digunakan sebagai sampel penelitian untuk menentukan sebuah pola yang akan digunakan dalam aplikasi. Data training diambil dari data survei pengeluaran dan pendapatan ekonomi keluarga di Kabupaten Ngawi. Atribut-atribut yang digunakan meliputi padi-padian, umbi-umbian, ikan/udang/cumi/kerang, daging, telur dan susu, sayur-sayuran, kacang-kacangan, buah-buahan, minyak dan kelapa, bahan minuman, bumbu-bumbuan, konsumsi lainnya, makanan dan minuman jadi, rokok, perumahan dan fasilitas rumah tangga, aneka barang dan jasa, pakaian, alas kaki dan tutup kepala, barang tahan lama, pajak, pungutan dan asuransi, keperluan pesta/upacara/kenduri, pendapatan dan tingkat kesejahteraan.

### 2.1.1.2 Data Testing

Data *testing* merupakan data yang belum terjadi yang nantinya digunakan untuk menguji pola dari data *training*. Data ini diambil dari data pengeluaran dan pendapatan ekonomi *user* yang telah diinputkan dalam aplikasi. Atribut-atribut yang digunakan meliputi padi-padian, umbi-umbian, ikan/udang/cumi/kerang, daging, telur dan susu, sayur-sayuran, kacang-kacangan, buah-buahan, minyak dan kelapa, bahan minuman, bumbu-bumbuan, konsumsi lainnya, makanan dan minuman jadi, rokok, perumahan dan fasilitas rumah tangga, aneka barang dan jasa, pakaian, alas kaki dan tutup kepala, barang tahan lama, pajak, pungutan dan asuransi, keperluan pesta/upacara/kenduri dan pendapatan.

## 2.2 Kebutuhan Data

Tahap ini merupakan tahap kebutuhan data yang digunakan dalam proses data mining untuk membantu pengembangan sistem. Pada tabel 2 menunjukkan atribut-atribut data penelitian.

Tabel 2. Atribut penelitian

Tingkat Kesejahteraan Keluarga	Y
Padi-padian	X1
Umbi-umbian	X2
Ikan/Udang/Cumi/Kerang	X3
Daging	X4
Telur dan Susu	X5
Sayur-sayuran	X6
Kacang-kacangan	X7
Buah-buahan	X8
Minyak dan Kelapa	X9
Bahan Minuman	X10
Bumbu-bumbuan	X11
Konsumsi Lainnya	X12
Makanan dan Minuman Jadi	X13
Rokok	X14
Perumahan dan Fasilitas Keluarga (Sebulan)	X15

Perumahan dan Fasilitas Keluarga (Setahun)	X16
Aneka barang dan Jasa (Sebulan)	X17
Aneka barang dan Jasa (Setahun)	X18
Pakaian, Alas kaki dan Tutup Kepala	X19
Barang Tahan Lama	X20
Pajak, Pungutan dan Asuransi	X21
Keperluan Pesta/ Upacara/Kenduri	X22
Pendapatan	X23

### 2.3 Pembersihan Data

Pembersihan data dilakukan untuk memperoleh data yang sesuai terhadap kebutuhan. Sehingga dari masing-masing atribut data pengeluaran dan pendapatan ekonomi keluarga tidak menimbulkan ketidakkonsistenan data atau *noise* saat melakukan uji aplikasi. Pembersihan data dilakukan dengan cara menyesuaikan atribut terhadap sistem.

### 2.4 Penggunaan Metode Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membangun sebuah pohon keputusan (*decision tree*) dari data yang telah ada. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang juga merupakan algoritma untuk membangun sebuah pohon keputusan. Algoritma C4.5 secara rekursif mengunjungi tiap simpul keputusan, memilih percabangan optimal, sampai tidak ada cabang lagi yang mungkin dihasilkan (Rahmayuni, 2014).

Untuk memilih atribut dengan nilai gain ratio terbesar menggunakan rumus Algoritma C4.5 (Ji dan Wei Dai, 2014) yang terdapat pada persamaan 2, 3, 4 dan 5.

$$\text{Entropi}(S) = - \sum_{j=1}^c p(S, j) \times \log p(S, j) \dots\dots\dots (2)$$

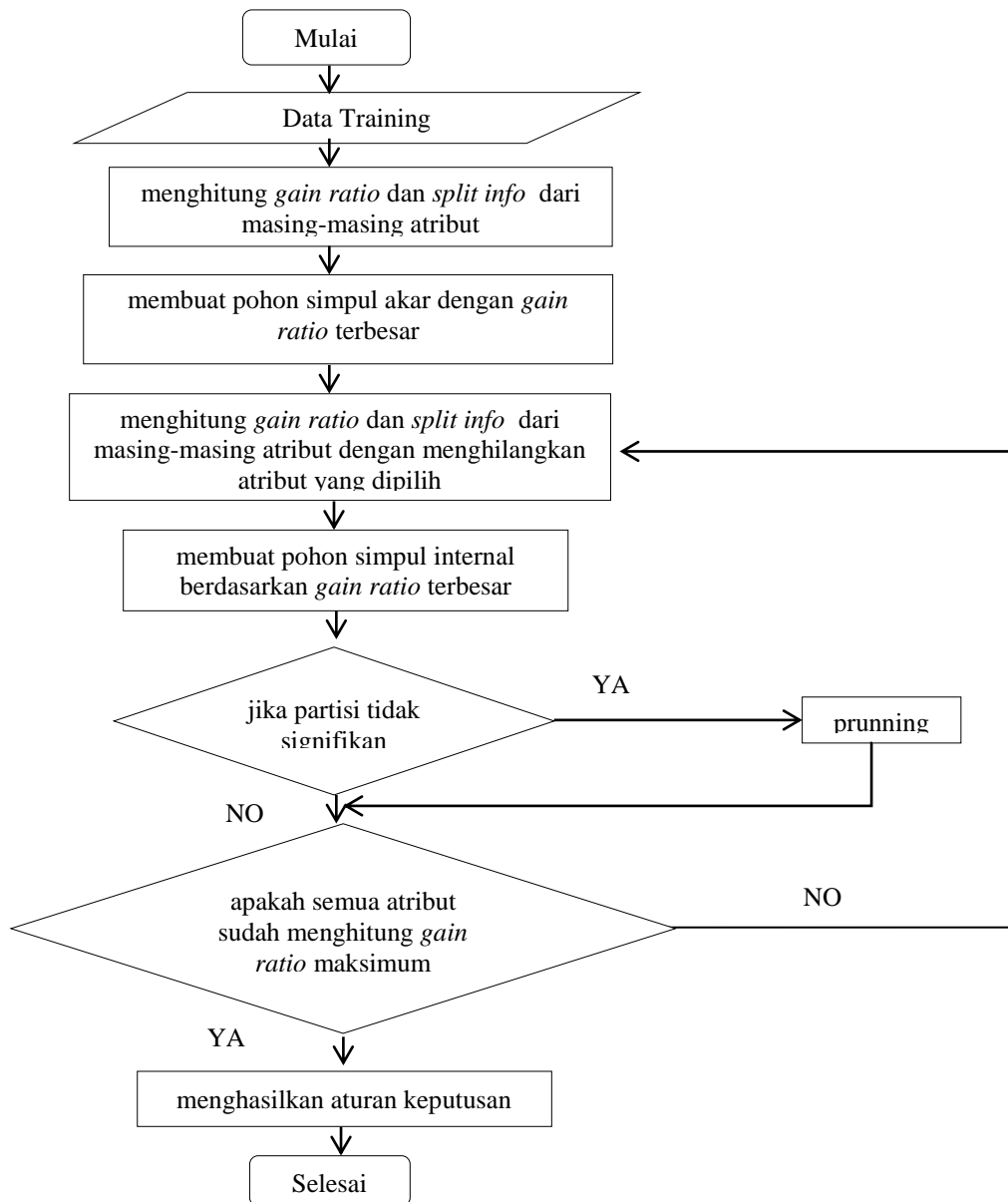
$$\text{Gain}(S, T) = \text{Entropi}(S) - \sum_{v \in \text{Values}(T_s)} \frac{T_{S,v}}{T_s} \text{entropi}(S_v) \dots\dots\dots (3)$$

$$\text{SplitInfo}(S, T) = \sum_{v \in \text{Values}(T_s)} \frac{T_{S,v}}{T_s} \times \log_2 \frac{T_{S,v}}{T_s} \dots\dots\dots (4)$$

$$\text{GainRatio}(S, T) = \frac{\text{Gain}(S, T)}{\text{SplitInfo}(S, T)} \dots\dots\dots (5)$$

Dimana algoritma C4.5 yang menghasilkan prediksi berdasarkan nilai *gain ratio* terbesar dari pola yang diketahui.

Amin,dkk. (2015) menyatakan Flowchart Algoritma C4.5 seperti gambar 2.



Gambar 2. Flowchart Algoritma C4.5

Tabel 3. Algoritma C4.5

#### Algoritma C4.5

- Data Training ( $X_1, X_2, X_3, X_4, X_5, Y$ )
- Menghitung *gain ratio* dan *split info* dari masing-masing atribut
- Membuat pohon simpul akar dengan *gain ratio* terbesar
- Menghitung *gain ratio* dan *split info* dari masing-masing atribut dengan menghilangkan atribut yang dipilih
- Membuat pohon simpul internal berdasarkan *gain ratio* terbesar
- Jika partisi tidak signifikan
  - Ya, melakukan proses pruning
- Apakah semua atribut sudah menghitung *gain ratio* maksimum
  - No, kembali pada tahap menghitung *gain ratio* dan *split info* dari

---

masing-masing atribut dengan menghilangkan atribut yang dipilih  
h. Menghasilkan aturan keputusan

---

### 3. HASIL DAN PEMBAHASAN

Pada tahap pengumpulan data dalam penelitian menghasilkan dua jenis data meliputi data *training* dan data *testing* dengan atribut seperti padi-padian, umbi-umbian, ikan/udang/cumi/kerang, daging, telur dan susu, sayur-sayuran, kacang-kacangan, buah-buahan, minyak dan kelapa, bahan minuman, bumbu-bumbuan, konsumsi lainnya, makanan dan minuman jadi, rokok, perumahan dan fasilitas rumah tangga, aneka barang dan jasa, pakaian, alas kaki dan tutup kepala, barang tahan lama, pajak, pungutan dan asuransi, keperluan pesta/upacara/kenduri dan pendapatan yang digunakan sebagai data pendukung aplikasi.

#### 3.1 Tampilan Program

Tampilan program pada halaman utama merupakan tampilan awal saat user membuka aplikasi kesejahteraan keluarga yaitu berupa informasi mengenai data kesejahteraan keluarga secara ekonomi. Pada halaman ini terdapat beberapa menu di kolom sebelah kiri yang dapat digunakan user diantaranya menu Home, Testing, Proses Mining, Perhitungan dan Pohon Keputusan. Sedangkan menu Login Admin merupakan menu dimana admin dapat melakukan login dengan cara memasukkan *username* dan *password*. Gambar 3 menunjukkan tampilan halaman utama dari aplikasi.



Gambar 3. Tampilan halaman utama

Pada menu testing, user dapat menginputkan data-data pengeluaran dan pendapatan ekonomi selama seminggu, sebulan dan setahun pada form testing yang akan tersimpan dalam *database*. User kemudian memilih tombol input untuk memproses data dengan algoritma C4.5. Gambar 4 menunjukkan tampilan form *testing* dan tabel 3 menunjukkan tabel data *testing* dari user.



## Testing

Input Data					
ID Data	:				
Pengeluaran Makanan		Pengeluaran Bukan Makanan		Pendapatan/ Penerimaan	
Jenis Pengeluaran	Seminggu (Rp)	Jenis Pengeluaran	Sebulan (Rp)	Setahun (Rp)	Pendapatan : Sebulan (Rp)
Padi-padian	< 11800	Perumahan dan Fasilitas Rumah Tangga	< 48300	< 579600	Pendapatan : < 790000
Umbi-umbian	< 2400	Aneka barang dan jasa	< 7000	< 174000	
Ikan/udang/cumi/kerang	< 5500	Pakaian, alas kaki dan tutup kepala		< 200000	
Daging	< 12000	Barang Tahan Lama		< 1500000	
Telur dan susu	< 8400	Pajak, Pungutan dan Asuransi		< 313500	
Sayur-sayuran	< 9400	Keperluan Pesta dan Upacara/Kenduri		< 4500000	
Kacang-kacangan	< 5600				
Buah-buahan	< 3800				
Minyak dan Kelapa	< 8900				
Bahan minuman	< 7600				
Bumbu-bumbuan	< 2100				
Konsumsi lainnya	< 3500				
Makanan dan minuman jadi	< 4500				
Rokok	< 12500				
Status Data	Data Testing				

Input

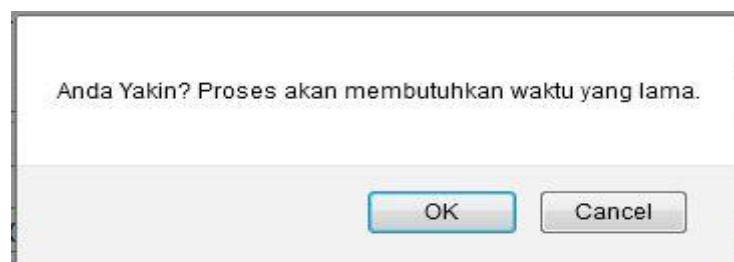
Gambar 4. Tampilan form *testing*

Data yang telah diinputkan oleh *user* menghasilkan output berupa tingkat kesejahteraan keluarga yang diperoleh dari pola yang terbentuk dalam data *training*. Data-data tersebut dikelompokkan dalam data *testing*. Tabel 3 menunjukkan tampilan data *testing*.

Tabel 4. Hasil data *testing*


NO	ID DATA	PADI-PADIAN	UMBI-UMBIA	IKAN/UDANG /CUMI/KERANG	DAGING	TELUR DAN SUSU	SAYUR-SAYURAN	KACANG-KACANGAN	BUAH-BUAHAN	MINYAK DAN KELAPA	BAHAN MINUMAN	BUMBU-BUMBUN	KONSUMSI LAINNYA	MAKANAN DAN MINUMAN JADI
1	001	<26200	<14400	<5500	<36000	<42000	<9400	<5600	<22800	<8900	<7600	<4300	<3500	<13500
2	002	<31000	<7200	<22000	<48000	<42000	<9400	<5600	<11400	<8900	<7600	<2100	<3500	<4500
3	003	<21400	<9600	<11000	<60000	<8400	<19600	<5600	<22800	<8900	<29200	<2100	<21000	<22500

Menu proses mining digunakan untuk memproses hasil input data *testing* dari *user* dengan menggunakan algoritma C4.5 untuk menghasilkan pohon keputusan. Gambar 5 menunjukkan tampilan proses mining.



Gambar 5. Tampilan konfigurasi proses

Hasil proses mining yang telah dilakukan, perhitungannya dapat dilihat dalam menu perhitungan. Gambar 6 menunjukkan hasil perhitungan algoritma C4.5.

<div>  <div> <b>APLIKASI KESEJAHTERAAN KELUARGA DENGAN METODE ALGORITMA C4.5</b>            KONSUMSI/PENGELUARAN MAKANAN DAN BUKAN MAKANAN,            DAN PENDAPATAN/PENERIMAAN RUMAH TANGGA         </div> </div>											
<div> <div> <a href="#">Home</a>  <a href="#">Testing</a>  <a href="#">C.45 » Proses Mining</a>  <a href="#">C.45 » Perhitungan</a>  <a href="#">C.45 » Pohon Keputusan</a>  <a href="#">Login Admin</a> </div> <div> <b>C45 » Perhitungan C4.5</b>            Ops: <a href="#">Hapus Semua Data</a> </div> </div>											
NO	ATRIBUT	GAIN RATIO MAX	ATRIBUT	NILAI ATRIBUT	JUMLAH KASUS TOTAL	JUMLAH KASUS TIDAK	JUMLAH KASUS SEJAHTERA	ENTROPY	INFORMATION GAIN	SPLIT INFO	GAIN RATIO
1	pendapatan	Total	Total	Total	399	88	311	0.7612			0
2	pendapatan	padi_padian	padi_padian	<11800	146	84	62	0.9836	0.3592	2.1739	0.1652
3	pendapatan	padi_padian	padi_padian	<16600	29	4	25	0.5788	0.3592	2.1739	0.1652
4	pendapatan	padi_padian	padi_padian	<21400	62	0	62	0	0.3592	2.1739	0.1652
5	pendapatan	padi_padian	padi_padian	<26200	129	0	129	0	0.3592	2.1739	0.1652
6	pendapatan	padi_padian	padi_padian	<31000	6	0	6	0	0.3592	2.1739	0.1652
7	pendapatan	padi_padian	padi_padian	<35800	19	0	19	0	0.3592	2.1739	0.1652
8	pendapatan	padi_padian	padi_padian	<40600	0	0	0	0	0.3592	2.1739	0.1652
9	pendapatan	padi_padian	padi_padian	<45400	7	0	7	0	0.3592	2.1739	0.1652
10	pendapatan	padi_padian	padi_padian	<50200	0	0	0	0	0.3592	2.1739	0.1652
11	pendapatan	padi_padian	padi_padian	>50200	1	0	1	0	0.3592	2.1739	0.1652
12	pendapatan	perumahan_dan_fasilitas_RT_sebulan	perumahan_dan_fasilitas_RT_sebulan	<48300	95	88	7	0.3795	0.6708	2.7888	0.2405
13	pendapatan	perumahan_dan_fasilitas_RT_sebulan	perumahan_dan_fasilitas_RT_sebulan	<84600	8	0	8	0	0.6708	2.7888	0.2405
14	pendapatan	perumahan_dan_fasilitas_RT_sebulan	perumahan_dan_fasilitas_RT_sebulan	<120900	14	0	14	0	0.6708	2.7888	0.2405
15	pendapatan	perumahan_dan_fasilitas_RT_sebulan	perumahan_dan_fasilitas_RT_sebulan	<157200	50	0	50	0	0.6708	2.7888	0.2405
16	pendapatan	perumahan_dan_fasilitas_RT_sebulan	perumahan_dan_fasilitas_RT_sebulan	<193500	74	0	74	0	0.6708	2.7888	0.2405

Gambar 6. Tampilan hasil perhitungan C4.5

Setelah data input *testing* dari *user* diolah menjadi pohon keputusan yang digunakan sebagai informasi untuk melihat atribut yang paling mempengaruhi kesejahteraan keluarga. Gambar 7 menunjukkan hasil pohon keputusan.

<div>  <div> <b>APLIKASI KESEJAHTERAAN KELUARGA DENGAN METODE ALGORITMA C4.5</b>            KONSUMSI/PENGELUARAN MAKANAN DAN BUKAN MAKANAN,            DAN PENDAPATAN/PENERIMAAN RUMAH TANGGA         </div> </div>											
<div> <div> <a href="#">Home</a>  <a href="#">Testing</a>  <a href="#">C.45 » Proses Mining</a>  <a href="#">C.45 » Perhitungan</a>  <a href="#">C.45 » Pohon Keputusan</a>  <a href="#">Login Admin</a> </div> <div> <b>C45 » Pohon Keputusan</b>            Ops: <a href="#">Hapus Semua Data</a> </div> </div>											
<b>Pohon Keputusan:</b> <pre> pendapatan = &lt;790000 (Tidak = 88, Sejahtera = 10) : ?   perumahan_dan_fasilitas_RT_sebulan = &lt;48300 (Tidak = 88, Sejahtera = 1) : ?     padi_padian = &lt;11800 (Tidak = 84, Sejahtera = 0) : Tidak       padi_padian = &lt;16600 (Tidak = 4, Sejahtera = 1) : Tidak     perumahan_dan_fasilitas_RT_sebulan = &lt;120900 (Tidak = 0, Sejahtera = 2) : Sejahtera   perumahan_dan_fasilitas_RT_sebulan = &lt;157200 (Tidak = 0, Sejahtera = 3) : Sejahtera   perumahan_dan_fasilitas_RT_sebulan = &lt;193500 (Tidak = 0, Sejahtera = 2) : Sejahtera   perumahan_dan_fasilitas_RT_sebulan = &lt;229800 (Tidak = 0, Sejahtera = 1) : Sejahtera   perumahan_dan_fasilitas_RT_sebulan = &lt;266100 (Tidak = 0, Sejahtera = 1) : Sejahtera pendapatan = &lt;1380000 (Tidak = 0, Sejahtera = 90) : Sejahtera pendapatan = &lt;1970000 (Tidak = 0, Sejahtera = 84) : Sejahtera pendapatan = &lt;2560000 (Tidak = 0, Sejahtera = 50) : Sejahtera pendapatan = &lt;3150000 (Tidak = 0, Sejahtera = 33) : Sejahtera pendapatan = &lt;3740000 (Tidak = 0, Sejahtera = 24) : Sejahtera pendapatan = &lt;4330000 (Tidak = 0, Sejahtera = 10) : Sejahtera pendapatan = &lt;4920000 (Tidak = 0, Sejahtera = 5) : Sejahtera pendapatan = &lt;5510000 (Tidak = 0, Sejahtera = 3) : Sejahtera pendapatan = &gt;5510000 (Tidak = 0, Sejahtera = 2) : Sejahtera           </pre>											
<b>Rule:</b> <pre> 1. if (pendapatan == &lt;790000 AND perumahan_dan_fasilitas_RT_sebulan == &lt;48300 AND padi_padian == &lt;11800) then Tidak (id = 3) 2. if (pendapatan == &lt;790000 AND perumahan_dan_fasilitas_RT_sebulan == &lt;48300 AND padi_padian == &lt;16600) then Tidak (id = 4) 3. if (pendapatan == &lt;790000 AND perumahan_dan_fasilitas_RT_sebulan == &lt;48300 AND padi_padian == &lt;21400) then (id = 5) 4. if (pendapatan == &lt;790000 AND perumahan_dan_fasilitas_RT_sebulan == &lt;48300 AND padi_padian == &lt;26200) then (id = 6) 5. if (pendapatan == &lt;790000 AND perumahan_dan_fasilitas_RT_sebulan == &lt;48300 AND padi_padian == &lt;31000) then (id = 7)           </pre>											

Gambar 7. Tampilan hasil pohon keputusan

Admin dapat menambah, mengedit nilai atribut data dan menghapus data *training* yang sudah ada. Gambar 8 menunjukkan form input data *training*.

### Training

**Menu: Data Pengeluaran | Cari Data**

Input Data						
ID Data	:					
Pengeluaran Makanan		Pengeluaran Bukan Makanan			Pendapatan/ Penerimaan	
Jenis Pengeluaran	Seminggu (Rp)	Jenis Pengeluaran	Sebulan (Rp)	Setahun (Rp)	Pendapatan	Sebulan (Rp)
Padi-padian	: < 11800 ▾	Perumahan dan Fasilitas Rumah Tangga	: < 48300 ▾	< 579600 ▾	Pendapatan	: < 790000 ▾
Umbi-umbian	: < 2400 ▾	Aneka barang dan jasa	: < 7000 ▾	< 174000 ▾		
Ikan/udang/cumi/kerang	: < 5500 ▾	Pakaian, alas kaki dan tutup kepala	:	< 200000 ▾		
Daging	: < 12000 ▾	Barang Tahan Lama	:	< 1500000 ▾		
Telur dan susu	: < 8400 ▾	Pajak, Pungutan dan Asuransi	:	< 313500 ▾		
Sayur-sayuran	: < 9400 ▾	Keperluan Pesta dan Upacara/Kenduri	:	< 4500000 ▾		
Kacang-kacangan	: < 5600 ▾					
Buah-buahan	: < 3800 ▾					
Minyak dan Kelapa	: < 8900 ▾					
Bahan minuman	: < 7600 ▾					
Bumbu-bumbuan	: < 2100 ▾					
Konsumsi lainnya	: < 3500 ▾					
Makanan dan minuman jadi	: < 4500 ▾					
Rokok	: < 12500 ▾					
class	: Tidak ▾					
Status Data	: Data Training ▾					

Gambar 8. Tampilan form input data *training*

## 3.2 Pembahasan

Hasil penelitian yang telah dilakukan dengan menggunakan metode algoritma C4.5 untuk memprediksi tingkat kesejahteraan keluarga di Kabupaten Ngawi sebagai berikut :

a. Menghitung entropi Ya dan Tidak.

$$\begin{aligned}
 \text{Entropi}(S) &= - \sum_{j=1}^c p(S,j) \times \log p(S,j) \\
 &= (-311/399) \times \text{LOG2}(311/399) + (-88/399) \times \text{LOG2}(88/399) \\
 &= 0,761173027\backslash
 \end{aligned}$$

b. Menghitung entropi padi-padian <11.800.

$$\begin{aligned}
 \text{Entropi}(S) &= - \sum_{j=1}^c p(S,j) \times \log p(S,j) \\
 &= (-62/146) \times \text{LOG2}(62/146) + (-84/146) \times \text{LOG2}(84/146) \\
 &= 0,983558567
 \end{aligned}$$

c. Menghitung entropi padi-padian <16.600.

$$\begin{aligned}\text{Entropi}(S) &= - \sum_{j=1}^c p(S,j) \times \log p(S,j) \\ &= (-29/146) \times \text{LOG2}(29/146) + (-25/146) \times \text{LOG2}(25/146) \\ &= 0,578794625\end{aligned}$$

d. Menghitung Gain information.

$$\begin{aligned}\text{Gain}(S, T) &= \text{Entropi}(S) - \sum_{v \in \text{Values}(T_s)} \frac{T_{S,v}}{T_s} \text{entropi}(S_v) \\ &= 0,761173027 ((146/399 \times 0,983558567) + (29/399 \times 0,578794625)) \\ &= 0,359206623\end{aligned}$$

e. Menghitung Split Info.

$$\begin{aligned}\text{SplitInfo}(S, T) &= \sum_{v \in \text{Values}(T_s)} \frac{T_{S,v}}{T_s} \times \log_2 \frac{T_{S,v}}{T_s} \\ &= (-146/399) \times \text{LOG2}(146/399) + (-29/399) \times \text{LOG2}(29/399) \\ &= 2,17388536\end{aligned}$$

f. Menghitung Gain Ratio.

$$\begin{aligned}\text{GainRatio}(S, T) &= \frac{\text{Gain}(S, T)}{\text{SplitInfo}(S, T)} \\ &= 0,359206623 + 2,17388536 \\ &= 0,165236909\end{aligned}$$

Hasil nilai *gain ratio* terbesar dari semua atribut dalam penelitian digunakan sebagai *root node* sebagai faktor yang mempengaruhi kesejahteraan keluarga.

### 3.3 Interpretasi Hasil

Berdasarkan contoh implementasi data *training* menggunakan aplikasi dapat diperoleh informasi mengenai faktor yang paling mempengaruhi kesejahteraan keluarga yaitu pendapatan kemudian diikuti faktor-faktor yang lain. Hasil tersebut dapat dilihat dari hasil pohon keputusan pada gambar 7 yang menunjukkan atribut pendapatan sebagai *root node*. Sehingga atribut tersebut dapat menjadi pertimbangan bagi keluarga dan pemerintah di masa mendatang.

## 4. PENUTUP

Berdasarkan hasil penelitian yang telah dilakukan, beberapa kesimpulan yang dapat diambil diantaranya :

- 1) Aplikasi kesejahteraan keluarga berbasis web dengan menggunakan metode Algoritma C4.5 telah berhasil dibangun untuk membantu keluarga dalam memprediksi tingkat kesejahteraan keluarga dan memberikan informasi kepada pemerintah tentang faktor-faktor yang mempengaruhi kesejahteraan berdasarkan atribut yang ada.



- 2) Faktor yang paling berpengaruh terhadap kesejahteraan keluarga berdasarkan *decision tree* adalah atribut yang menempati sebagai *root node* dengan nilai *gain ratio* terbesar. Dalam penelitian ini, pendapatan menempati sebagai *root node*, sehingga pendapatan merupakan atribut yang paling berpengaruh terhadap tingkat kesejahteraan keluarga di kabupaten Ngawi.

## DAFTAR PUSTAKA

- Amin, R. K. Indwiarti & Yuliant, S. 2015. *Implementation of Decision Tree Using C4.5 Algorithm in Decision Making of Loan Application by Debtor (Case Study: Bank Pasar of Yogyakarta Special Region)*. Available at : [http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=7231400&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs\\_all.jsp%3Farnumber%3D7231400](http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=7231400&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D7231400) [Accessed 27 Oktober 2015].
- BPS Kabupaten Ngawi. 2015. Available at : <http://ngawikab.bps.go.id/Subjek/view/id/23#subjekViewTab1|accordion-daftar-subjek1> [Accessed 27 Oktober 2015].
- BPS. 2015. *Profil Kemiskinan Di Indonesia Maret 2015*. Berita Resmi Statistik Badan Pusat Statistik. Available at : [http://bps.go.id/website/brs\\_ind/brsInd-20150915122517.pdf](http://bps.go.id/website/brs_ind/brsInd-20150915122517.pdf) [Accessed 27 Oktober 2015].
- Devi, J. C. 2014. *Binary Decision Tree Classification based on C4.5 and KNN Algorithm for Banking Application*. Available at : <http://www.periyaruniversity.ac.in/ijcii/issue/Vol4No2September2014/IJCII-4-2-145.pdf> [Accessed 28 Maret 2016].
- Ji, Wei dan Wei Dai. 2014. *A Map Reduce Implementation of C4.5 Decision Tree Algorithm*. Available at : [http://www.sersc.org/journals/IJDTA/vol7\\_no1/5.pdf](http://www.sersc.org/journals/IJDTA/vol7_no1/5.pdf) [Accessed 28 Maret 2016].
- Kamagi, David Hartanto & Hansun, Seng. 2014. *Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa*. Available at : <http://library.umn.ac.id/jurnal/public/uploads/papers/pdf/be20fbb34fbc7c5d2fe06044edf6d37e.pdf> [Accessed 28 Oktober 2015].
- Nugroho, Yusuf Sulistyo. 2014. *Modul Praktikum Data Mining Berdasarkan Kurikulum 2013*. Surakarta: Program Studi Teknik Informatika Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta.
- Nugroho, Yusuf Sulistyo. & Haryati, Syarifah N. 2015. *Klasifikasi dan Klustering Penjurusan Siswa SMA Negeri 3 Boyolali*. Jurnal Khazanah Informatika Vol. I No. 1 Desember 2015 Available at : <http://journals.ums.ac.id/index.php/khif/article/view/1175/1012> [Accessed 24 Desember 2015]
- Nugroho, Yusuf Sulistyo dan Setyawan. 2014. *Klasifikasi Masa Studi Mahasiswa Fakultas Komunikasi dan Informatika*. Available at : [https://publikasiilmiah.ums.ac.id/bitstream/handle/11617/4487/1\\_Yusuf%20Sulistyo%20Nugroho.pdf?sequence=1](https://publikasiilmiah.ums.ac.id/bitstream/handle/11617/4487/1_Yusuf%20Sulistyo%20Nugroho.pdf?sequence=1) [Accessed 24 Desember 2015]
- Rahmayuni, Indri. 2014. *Perbandingan Performansi Algoritma C4.5 Dan Cart Dalam Klasifikasi Data Nilai Mahasiswa Prodi Teknik Komputer Politeknik Negeri Padang*. Available at : <https://ejournal.itp.ac.id/index.php/tinformatika/article/download/186/182> [Accessed 02 Januari 2016]